Med-VRAgent: A Framework for Medical Visual Reasoning-Enhanced Agents

Anonymous Submission

Abstract

Vision-language models (VLMs) achieve promising results in medical reasoning but struggle with hallucinations, vague descriptions, Inconsistent logic and poor localization. To address this, we propose a agent framework named Medical Visual Reasoning Agent (**Med-VRAgent**). The approach is based on Visual Guidance and Self-Reward paradigms and Monte Carlo Tree Search (MCTS). By combining the Visual Guidance with tree search, Med-VRAgent improves the medical visual reasoning capabilities of VLMs. We use the trajectories collected by Med-RAgent as feedback to further improve the performance by fine-tuning the VLMs with the proximal policy optimization (PPO) objective. Experiments on multiple medical VQA benchmarks demonstrate that our method outperforms existing approaches.

1 Introduction

Visual Language Models (VLMs) enable context-aware medical reasoning and have shown strong performance in tasks like radiology report generation (Hartsock and Rasool, 2024; Tanno et al., 2025; Li et al., 2023c). However, they remain prone to hallucinations, where outputs deviate from the visual input—posing risks in clinical settings (Chen et al., 2025; Jin et al., 2024). This issue is exacerbated by the factual unreliability of underlying large language models (LLMs) (Huang et al., 2023; Zhu et al., 2024b; Pal et al., 2023), highlighting the urgent need for effective mitigation strategies (Kim et al., 2025; Huang et al., 2025).

Researchers also have explored several enhancements, the Chain-of-Thought (CoT) has become a popular approach to enhance the logical reasoning capability (Wei et al., 2022). Visual prompting—using region-specific cues —have been shown to improve model performance in fields such as radiology and pathology where precise localization is required. (Denner et al., 2025).

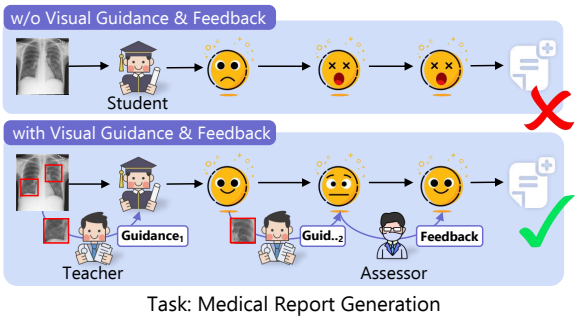

Figure 1: Top: A student struggles, feeling confused and making mistakes. Bottom: With guidance, the student overcomes the confusion and successfully completes the task.

Plan-then-Generate decouple reasoning into structured planning followed by execution (Zhou et al., 2022). Self-enhancement mechanisms, such as self-reflection, self-correction, and self-critique, and external feedback systems enable models to revise their own reasoning (Madaan et al., 2023; Liu et al., 2023b). Additionally, Retrieval-Augmented Generation (RAG) incorporates external knowledge to support the reasoning process (Lewis et al., 2020).

While the above approaches are effective, some key challenges remain. **(1)** In high-stakes domains like radiology and pathology, VLMs often lack fine-grained image-text alignment, producing overly generic descriptions that miss critical local details, spatial structures, and abnormal patterns. (Zhang et al., 2023a; Yamamoto et al., 2023). **(2)** Although complex medical prompting strategies have been proposed to address this issue, they are often domain-specific, labor-intensive, and require expert knowledge. (Boiko et al., 2023; Yu et al., 2023). **(3)** Furthermore, current models, even with visual prompting, focus on a single ROI and struggle to integrate overall medical image structure and spatially distributed lesions, limiting performance in cases with high spatial complexity. (Wang et al., 2023; Cheng et al., 2023). **(4)** Current frameworks offer limited feedback, usually evaluating only the final output, making error detection and correction

during reasoning difficult. **(5)** Finally, retrieval enhancement methods often introduce irrelevant or noisy information, potentially distorting clinical reasoning. (Gao et al., 2023; Ji et al., 2023).

We propose a multimodal agent framework **Med-RAgent**, to tackle challenges like error propagation, suboptimal planning, limited feedback, and the fragility of retrieval-based methods. Med-RAgent consists of three core modules—**Teacher**, **Student**, and **Assessor**—and two key components: a **Visual Extraction Module**, and a **Retrieval-Augmented Reflection (RAR)**. The Visual Extraction Module identifies Regions of interest (ROIs) in medical images and uses Visual Token Edit to improve the agent's regional perception. The Teacher provides ROI-specific visual guidance. The Student generate diagnostic outputs with ROI and teacher's guidance. The Assessor offers fine-grained feedback for iterative refinement. We use RAR module to enhance factual grounding by incorporating external medical knowledge and introduce Monte Carlo Tree Search (MCTS) to explore high-quality reasoning paths using an adaptive expansion strategy while better balancing performance and efficiency. Our framework only needs to be trained once for both the teacher and the assessor, which can achieve good generalization ability and save computational resources.

Results across three benchmarks confirm the superior performance of Med-VRAgent, achieving new state-of-the-art (SOTA) results. It outperforms reasoning baselines (Visual CoT) on GMAI (Table 3), surpasses retrieval-augmented methods on IU-Xray (Demner-Fushman et al., 2015) (Table 4), and exceeds advanced fine-tuning strategies like MMedPO on VQA-RAD (Lau et al., 2020) and MIMIC-CXR (Johnson et al., 2019) (Table 2). These results highlight the effectiveness of our visual guidance-based medical multimodal agent framework.

In summary, our contributions are as follows:

- We propose a **Teacher-Student-Evaluator** framework for medical visual reasoning based on **Visual Guidance** and **Feedback**.
- We use **Visual Extraction** and **Visual Token Edit** to improve the visual capabilities of multimodal agents.
- We develop a **Retrieval-Augmented Reflection** module to further boost agent reasoning via External knowledge.
- Extensive experiments on multiple medical multimodal benchmarks demonstrate that our framework achieves SOTA performance.

2 Related Work

2.1 Foundation Large models

Large Language Models (LLMs) like GPT-3 (Brown et al., 2020), PaLM (Chowdhery et al., 2022), and LLaMA (Touvron et al., 2023) have shown strong capabilities in reasoning, generation, and understanding across natural language tasks, excelling in few-shot learning, in-context reasoning, and text generation. These models are central to the development of multi-modal systems. VLMs have demonstrated remarkable generalization across cross-modal tasks such as image captioning, retrieval, and visual question answering (VQA). Early models like CLIP (Radford et al., 2021) and Flamingo (Alayrac et al., 2022) use large-scale image-text pairs for contrastive or retrieval-based learning. Recent models like BLIP-2 (Li et al., 2023b) and MiniGPT-4 (Zhu et al., 2023) integrate LLMs with visual encoders to enhance reasoning and support open-ended question answering. These advances in Foundation Large Models (FLMs) lay the foundation for tasks that require deep cross-modal understanding.

2.2 Multi-step Reasoning in FLMs

Reasoning in Foundation Large Models (FLMs) has advanced with frameworks enhancing multi-step inference and decision-making. CoT (Wei et al., 2023) enables intermediate reasoning steps, improving performance on complex tasks. ToT (Yao et al., 2023) explores multiple reasoning paths using tree search strategies, boosting decision-making. The ReAct framework (Yao et al., 2022) combines reasoning with environment interaction, improving tool-augmented tasks. In multimodal reasoning, Visual Chain-of-Thought (Rose et al., 2024) extends CoT by integrating visual grounding to bridge logical gaps. The Reinforced Ranker-Reader (R^3) architecture (Zhang et al., 2023c) improves open-domain question answering by combining a ranker and reader with reinforcement learning, optimizing accuracy over retrieved documents.

2.3 Medical-Specific Reasoning Frameworks

Medical-Specific Reasoning has advanced with specialized frameworks to enhance LLMs' clinical reasoning. MedAgents (Tang et al., 2023) creates a multi-agent system where LLM-based

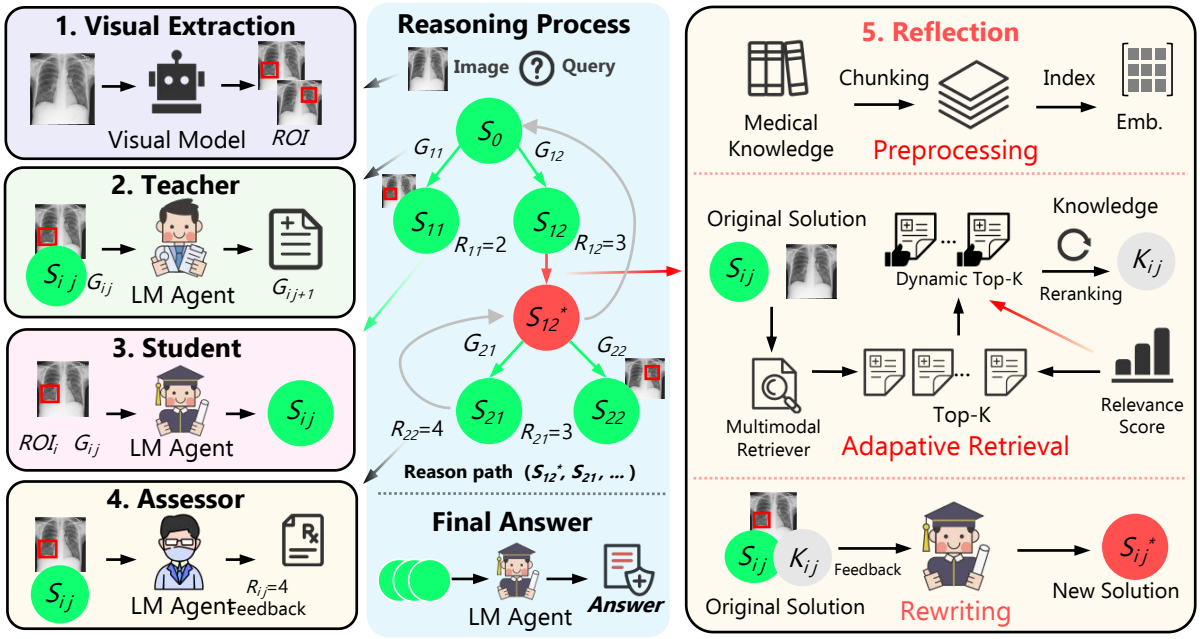

Figure 2: Overview of the **Med-VRAgent** framework. The system uses MCTS to generate solutions $\mathcal{S}_{ij}$ based on Regions of Interest $\mathcal{ROI}_{ij}$, visual guides $\mathcal{G}_{ij}$, rewards $\mathcal{R}_{ij}$, and external knowledge $\mathcal{K}_{ij}$.$\mathcal{S}^*_{ij}$ is the solution after reflection.

medical experts collaborate on diagnostic tasks, improving zero-shot reasoning. MedReason (Wu et al., 2025) aligns LLM reasoning with medical graphs, enhancing decision-making accuracy and interpretability. FineMedLM-o1 (Yu et al., 2025) uses supervised fine-tuning and test-time training on curated dialogues for complex tasks like differential diagnosis. DeepSeek R1 (Moell et al., 2025) benchmarks LLM outputs against expert behavior, revealing both advanced reasoning and domain-specific biases. These models highlight the value of tailored frameworks and medical knowledge in improving LLM clinical reasoning.

3 Methodology

To enhance medical visual reasoning, we propose Med-VRAgent, a novel reasoning scheme. It combines a Visual guidance and Reward-Feedback Paradigm in a search algorithm to optimize reasoning paths.

3.1 Med-VRAgent Reasoning Process

Fig 2 illustrates the Med-VRAgent process. We model the agent reasoning process as a tree search, where each node $\mathcal{S}_{ij}$ represents a state defined as:

$$\mathcal{S}_{ij} = [\mathcal{Q}, \mathcal{I}, \mathcal{G}_{ij}, \mathcal{A}_{ij}, \mathcal{R}_{ij}, \mathcal{F}_{ij}, \mathcal{A}^*_{ij}, \mathcal{O}_{ij}, \mathcal{ROI}_i] \quad (1)$$

where $\mathcal{Q}$ is the query, $\mathcal{I}$ is the medical image, $\mathcal{G}_{ij}$ is the visual guidance, $\mathcal{A}_{ij}$ is the current step answer, $\mathcal{R}_{ij}$ is the reward, $\mathcal{F}_{ij}$ is feedback, $\mathcal{A}^*_{ij}$ is the answer after reflection, $\mathcal{O}_{ij}$ represents the observation information, including all ancestor and sibling node guidance and answers, and $\mathcal{ROI}_i$ is the visual ROI.

Given an image $\mathcal{I}$ and query $\mathcal{Q}$, the goal is for Student $\mathcal{S}_{\text{model}}$ to generate step-by-step reasoning using ROIs $\mathcal{ROI}_i$ from Vision Extraction $\mathcal{V}_{\text{model}}$ and visual guidance $\mathcal{G}_{ij}$ from Teacher $\mathcal{T}^{\theta}_{\text{model}}$. Asserror $\mathcal{A}^{\theta}_{\text{model}}$ evaluates guidance and answers, providing reward $\mathcal{R}_{ij}$ and feedback $\mathcal{F}_{ij}$. If answer quality is low, the reflection module uses external knowledge $\mathcal{K}$ from retriever $\mathcal{R}_{\text{model}}$ to refine it. MCTS searches for the optimal reasoning path for answering $\mathcal{Q}$.

3.2 Visual Extraction Module

Visual ROIs Extraction We use a lightweight VLM to extract medical entities E relevant to the question and image. Following MedVP (Zhu et al., 2025), we adopt a fine-tuned Grounding DINO (Liu et al., 2024) as the visual extractor. Grounding DINO is an open-vocabulary detector that localizes entities from image I and text prompts $E = \{e_1, e_2, \ldots, e_N\}$.

$$ROI = \{ROI_i\}_{i=1}^{N} = \text{G-DI}(I, E), \quad ROI_i = (b_i, s_i, l_i) \quad (2)$$

ROI is the set of extracted regions, with each $ROI_i = (b_i, s_i, l_i)$ representing the bounding box, confidence score, and matched entity label.

Visual Token Edit To refine the Agent's focus on a ROI without retraining, we apply Visual Token Edit (VTE), a single edit to visual tokens in the first ($K \leq 3$) self-attention layers. For each patch embedding $\mathbf{v}_i \in \mathbb{R}^d$ and binary ROI mask $m_i \in \{0, 1\}$, we replace:

$$\mathbf{v}_i \longrightarrow \mathbf{v}_i^\star = \mathbf{v}_i + \beta\, m_i\, \mathbf{b} \tag{3}$$

where $\mathbf{b}$ is a fixed direction (e.g., $\mathbf{1}$ or $\mathbf{v}_i$). Because the key and value projections are linear, Eq. (3) increases the ℓ_2 norm of ROI tokens and thus raises their soft-max attention weights *indirectly*, concentrating information flow on the referenced region while keeping background tokens intact.

The gain $\beta > 0$ is chosen on-the-fly to prevent over- or under-boosting:

$$\beta = s_i\, \kappa\, \phi\Big(\frac{\bar{a}_{\text{bg}}}{\bar{a}_{\text{ROI}}} - 1\Big), \qquad \kappa \in [0, 1], \tag{4}$$

where $\bar{a}_{\text{ROI}}$ and $\bar{a}_{\text{bg}}$ are the average pre-softmax attention logits of ROI and background patches obtained from a provisional forward pass, and s_i is the detector confidence for the ROI, $\phi(\cdot)$ is any element-wise activation that is non-negative and monotonically non-decreasing. When the model already attends to the ROI ($\bar{a}_{\text{ROI}} \geq \bar{a}_{\text{bg}}$), Eq. (4) yields $\beta = 0$, leaving the representation unchanged. Setting $\kappa = 0$ disables VTE entirely, making the mechanism safe,computationally negligible, and fully reversible.

3.3 Teacher-Student-Assessor Mechanism

Teacher Agent. In natural language tasks, the exponential growth of tag combinations severely limits vanilla MCTS. To improve efficiency, we incorporate a prompt-driven Teacher $\mathcal{T}^\theta_{\text{model}}$ that expands the policy space via heuristics. See the appendix 6 for prompt. At each node, $\mathcal{T}^\theta_{\text{model}}$ gathers prior guidance–answer pairs $(\mathcal{G}_{1..i}, \mathcal{A}_{1..i})$ and feedback $\mathcal{F}$, then generates the next-step guidance:

$$\mathcal{G}_{ij+1} = \mathcal{T}_{\text{model}}(\mathcal{ROI}_i, \mathcal{G}_{i1..j}, \mathcal{A}_{i1..j}, \mathcal{F}_i) \tag{5}$$

Student Agent. The Student $\mathcal{S}_{\text{model}}$ leverages a vision-language backbone to perform step-wise reasoning. At each stage of problem, it receives the Teacher $\mathcal{T}^\theta_{\text{model}}$-generated guidance $\mathcal{G}_{ij}$ and the corresponding image $\mathcal{ROI}_i$, and produces an intermediate answer $\mathcal{A}_{ij}$. After search, the best reasoning path selected by MCTS is used to compose the final answer. This process is formally defined as:

$$\mathcal{A}_{ij} = \mathcal{S}_{\text{model}}(\mathcal{ROI}_i,\, \mathcal{G}_{ij}), \tag{6}$$

Assessor Agent. In the MCTS, it is essential to quantitatively evaluate each reasoning step and provide high-quality feedback to guide the search process. To this end, we adopt a *LLM-as-a-Judge* (Gu et al., 2025) approach, we introduce an Assessor model $\mathcal{A}^\theta_{\text{model}}$, implemented using a VLM, and grounded in the *Self-Rewarding* paradigm (Yuan et al., 2025) The Assessor $\mathcal{A}^\theta_{\text{model}}$ employs a 5-point scoring system to evaluate task progress, where the score reflects both the quality and contribution of each intermediate answer. The Assessor $\mathcal{A}^\theta_{\text{model}}$ receives the image ROI $\mathcal{ROI}_i$, the current guidance $\mathcal{G}_{ij}$, and the student's answer $\mathcal{A}_{ij}$. It then produces both a descriptive feedback $\mathcal{F}_{ij}$ and a quantitative rating $\mathcal{R}_{ij}$, See Appendix 7 for prompt. The process is formalized as:

$$\mathcal{F}_{ij},\, \mathcal{R}_{ij} = \mathcal{A}_{\text{model}}(\mathcal{ROI}_i,\, \mathcal{G}_{ij},\, \mathcal{A}_{ij}). \tag{7}$$

3.4 Retrieval-Augmented Reflection

The reflection phase is designed to enhance ROI analysis tasks that the Student $\mathcal{S}_{\text{model}}$ fails to complete under Teacher $\mathcal{T}^\theta_{\text{model}}$ guidance. We use IU-Xray, MIMIC-CXR, VQA-RAD and other datasets as knowledge sources. The reflection process consists of two stages:

Retrieval Phase. We adopt the domain-aware retriever from MMed-RAG (Xia et al., 2025), which uses ResNet-50 (He et al., 2015) and BioClinical-BERT (Alsentzer et al., 2019) as the image and text encoders, respectively. During reflection, the retriever takes as input the guidance $\mathcal{G}_{ij}$, image $\mathcal{I}$, and answer $\mathcal{A}_{ij}$. It first retrieves a Top-$\mathcal{K}$ candidate set $\mathcal{K}_1$ from the external knowledge base using FAISS (Johnson et al., 2017). A cross-attention-based relevance scoring model cross-encoder/ms-marco-MiniLM-L-6-v2 (Reimers and Gurevych, 2019) then refines these candidates into a subset $\mathcal{K}_2$, which is finally reranked to produce the final knowledge set $\mathcal{K}_{ij}$. This multi-stage knowledge retrieval process is formally expressed as:

$$\mathcal{K}_{ij} = \text{Rerank}\Big(\text{Relevance}\big(\text{RetrieveTop-}K(\mathcal{I}, \mathcal{G}_{ij}, \mathcal{A}_{ij})\big)\Big) \tag{8}$$

Rewriting Phase. When reflection is needed, the student $\mathcal{S}_{\text{model}}$ receives the original answer $\mathcal{A}_{ij}$, guidance $\mathcal{G}_{ij}$, the input $\mathcal{ROI}$, feedback $\mathcal{F}_{ij}$, and retrieved knowledge $\mathcal{K}_{ij}$. It then synthesizes these inputs to produce a refined answer $\mathcal{A}^*_{ij}$. This rewriting process can be formalized as:

$$\mathcal{A}^*_{ij} = \mathcal{S}_{\text{model}}(\mathcal{ROI}_i, \mathcal{G}_{ij}, \mathcal{A}_{ij}, \mathcal{F}_{ij}, \mathcal{K}_{ij}) \tag{9}$$

3.5 Monte Carlo Tree Search Process

Monte Carlo Tree Search (MCTS) operates through four main phases—selection, expansion, evaluation, and backpropagation—repeating until satisfactory reasoning results are produced or computational limits are reached. In the **Selection** phase, the algorithm starts at the root node (initial state $\mathcal{S}_0$) and recursively selects child nodes using the Upper Confidence Bound (UCB) formula, which balances exploration and exploitation:

$$UCB(s) = \frac{R(s)}{N(s)} + C \cdot \sqrt{\frac{2 \cdot \ln N(p)}{N(s)}} \quad (10)$$

where $R(s)$ is the reward, $N(s)$ the visit count of node s, $N(p)$ the visit count of its parent p, and C is a constant. The **Expansion** phase involves selecting an unprocessed ROI along the current path and expanding it by sampling $\mathcal{N}$ guidance suggestions from the Teacher $\mathcal{T}^{\theta}_{\text{model}}$. This step incorporates a heuristics mechanism, where feedback from Assessor $\mathcal{A}^{\theta}_{\text{model}}$ and all observations—including guidance, answer from ancestor and sibling nodes are provided to the Teacher $\mathcal{T}^{\theta}_{\text{model}}$. In the **Evaluation** phase, each new child node is assessed using feedback from the Assessor $\mathcal{A}^{\theta}_{\text{model}}$. Finally, in the **Backpropagation** phase, the reward $\mathcal{R}(s')$ is used to update the average reward and visit counts for node $\mathcal{S}'$ and its ancestors.

To improve search performance and efficiency in MCTS, we apply some strategy.

Early Stopping. Expansion is terminated when the node score exceeds 4 or when KL divergence and semantic similarity suggest the Student $\mathcal{S}_{\text{model}}$ and Teacher $\mathcal{T}^{\theta}_{\text{model}}$ outputs align with the previous node. This allows the agent to shift to other ROIs.

Alpha-Beta Pruning. During selection and expansion, Alpha (min guaranteed by maximization) and Beta (max guaranteed by minimization) bounds are maintained. Subtrees are pruned when node scores fall outside this range, avoiding unnecessary evaluations.

Reflection. If early stopping is triggered repeatedly or the expansion limit is reached without achieving a score of 4, the reflection module is activated. In this case, the Student $\mathcal{S}_{\text{model}}$ retrieves external knowledge to continue reasoning.

3.6 Training Strategy and Optimization

To enhance the Teacher $\mathcal{T}^{\theta}_{\text{model}}$ and Assessor $\mathcal{A}^{\theta}_{\text{model}}$, we fine-tune both VLMs using proximal policy optimization (PPO) with feedback trajectories collected by Med-VRAgent. PPO optimizes the policy by maximizing expected rewards while constraining updates to avoid performance degradation. The objective is:

$$\mathcal{L}_{\text{PPO}}(\theta) = \mathbb{E}\left[\min\left(r_\theta \hat{A}_t,\ \text{clip}(r_\theta, 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right] \quad (11)$$

where

$$r_\theta = \frac{\pi_\theta(A_{1..i}|O_{1..i})}{\pi_{\theta_{\text{old}}}(A_{1..i}|O_{1..i})} \quad (12)$$

Here, $A_{1..i}$ and $O_{1..i}$ denote sampled actions (guidance) and observations, respectively, while $\hat{A}_t$ is the advantage estimate and ϵ is the clipping threshold. We collect trajectories

$$\mathcal{T}_{\text{Med-VRAgent}} = (A_{1..i}, O_{1..i}, R_{1..i}) \quad (13)$$

from Med-VRAgent to estimate advantages and update the policy parameters θ. The clipping in $\mathcal{L}_{\text{PPO}}(\theta)$ ensures conservative, stable updates.

4 Experiments

4.1 Experimental Datasets

Dataset	Modality	Size	Task Type
IU-Xray	X-ray	590	Report Generation
MIMIC-CXR	Chest X-ray	500	Report Generation
VQA-RAD	X-ray, CT	451	Visual Question Answering
GMAI-MMbench	38 modalities	4 task	Visual Question Answering

Table 1: The medical visual dataset used in this experiment

We evaluate Med-VRAgent on various medical visual-linguistic datasets covering report generation and VQA tasks. As shown in Table 1, for report generation, we use the IU-Xray (Demner-Fushman et al., 2015) dataset containing 590 test samples and the MIMIC-CXR (Johnson et al., 2019) dataset test with 500 test samples. For VQA, we use VQA-RAD containing test 451 QA pairs based on X-rays and CT images and GMAI-MMbench (Chen et al., 2024b) we use 4 clinical tasks.

For the Med-VQA task, for open questions, we report recall in the Open column. For closed questions, we report precision in the Closed column. For the report generation task, we use BLEU (Papineni et al., 2002) Score, ROUGE-L (Lin, 2004), and METEOR as metrics (Banerjee and Lavie, 2005). BLEU score represents the average of BLEU-1/2/3/4.

4.2 Compared Methods

We evaluate the performance of various methods across different approaches.

For training methods, we employ the LLaVA-Med (Li et al., 2023a) model and assess its performance on the VQA-RAD and MIMIC-CXR datasets. The training approaches compared include SFT, Self-Rewarding (Yuan et al., 2025), Direct Preference Optimization (DPO) (Rafailov et al., 2024), STLLaVA-Med, and MMedPO (Zhu et al., 2024a).

For reasoning methods, we use the DeepSeek-VL-7B (Lu et al., 2024) and MiniCPM-V2 (Yao et al., 2024) models, evaluating their performance on the GMAI-MMbench. The reasoning approaches compared include CoT (Wei et al., 2023), ToT (Yao et al., 2023), and Visual CoT (Shao et al., 2024).

Finally, for Decoding-based and Retrieval-Augmented methods, we use the LLaVA-Med v1.5 model and evaluate its performance on the IU-Xray dataset. The Decoding-based methods include Greedy Decoding, BeamSearch (Xie et al., 2023), DoLa (Chuang et al., 2024), OPERA (Huang et al., 2024), VCD (Leng et al., 2023). The RAG approaches compared include MedDr (He et al., 2024), FactMM-RAG (Zhang et al., 2023b), RULE (Liu et al., 2023a), and MMed-RAG (Chen et al., 2024a). Please see the appendix 6 for details.

4.3 Model Implementation

We applied Med-VRAgent to LLaVA-Med v1.5, DeepSeek-VL-7B, and MiniCPM-V2. To ensure fair comparison, we follow the same experimental settings as prior work, using a decoding temperature of 0.7. We use DeepSeek-VL-7B as the Teacher $\mathcal{T}^{\theta}_{\text{model}}$ and Assessor $\mathcal{A}^{\theta}_{\text{model}}$ and perform PPO fine-tuning.

For PPO fine-tuning, We follow the official training scripts and use the ‘peft‘ and ‘trl‘ Python packages to implement LoRA and PPO. The fine-tuning process is completed within 7–8 hours on 4 Nvidia A6000 GPUs. The ‘lora_target_modules‘ are set to ["q_proj", "v_proj"], with lora_r set to 16, lora_alpha set to 32, and lora_dropout set to 0.05. The micro_batch_size is 1, the batch_size is 8, and num_epochs is 1. For optimization, we set the learning_rate to 1.41e-5, the reward baseline to 3.75, and the random seed to 0.

Methods	VQA-RAD		MIMIC-CXR		
	Open	Closed	BLEU	ROUGE-L	METEOR
LLaVA-Med v1.5	29.24	63.97	2.76	10.31	7.71
SFT	31.38	64.26	3.10	10.21	8.75
Self-Rewarding	32.69	65.89	3.03	10.05	8.77
DPO	32.88	64.33	3.17	10.38	9.10
STLLaVA-Med	33.72	64.70	3.05	10.12	8.98
MMedPO	34.03	67.64	3.71	13.22	**10.20**
Med-VRAgent (Ours)	**35.70**	**68.72**	**3.90**	**13.53**	9.58

Table 2: Comparison of Med-VRAgent with fine-tuning methods, including SFT, Self-Rewarding, DPO, STLLaVA-Med, and MMedPO, evaluated on VQA-RAD (Open/Closed Accuracy) and MIMIC-CXR (BLEU, ROUGE-L, METEOR) datasets, based on LLaVA-Med v1.5. The best result for each model is bolded.

Methods	AR	BVR	B	CR	Average
DeepSeek-VL-7B	38.43	47.03	42.31	37.03	41.20
CoT	39.24	46.60	43.26	38.18	41.57
ToT	40.23	46.07	44.42	39.58	42.08
Visual CoT	41.57	46.76	44.13	41.59	43.51
Med-VRAgent (Ours)	**44.81**	**51.82**	**47.52**	**42.79**	**46.74**
MiniCPM-V2	40.74	43.01	36.46	37.57	39.45
CoT	41.69	43.90	37.69	38.74	40.51
ToT	42.14	44.32	38.27	39.29	41.01
Visual CoT	43.20	44.70	39.12	41.28	42.08
Med-VRAgent (Ours)	**44.81**	**47.32**	**40.18**	**41.34**	**43.41**

Table 3: Comparison of Med-VRAgent with reasoning methods, including CoT, Tree-of-Thought (ToT), and Visual CoT ,evaluated on the GMAI (Accuracy) dataset, based on DeepSeek-VL-7B and MiniCPM-V2. GMAI include AR (Attribute Recognition), BVR (Blood Vessels Recognition), B (Bone), and CR (Cell Recognition). The best result for each model is bolded, and average values are in blue.

4.4 Overall Performance

Evaluating Training Strategy. As shown in Table 2, we evaluated Med-VRAgent on medical VQA tasks using the LLaVA-Med v1.5 model, comparing it with five baselines: Zero-shot, SFT, Self-Rewarding, DPO, and STLLaVA-Med. On VQA-RAD, Med-VRAgent achieved 35.70 (open) and 68.72 (closed); on MIMIC-CXR, it scored 3.90 (BLEU), 13.53 (ROUGE-L), and 9.58 (METEOR), outperforming other methods in generalization and generation quality.

Evaluating Reasoning Strategy. As shown in Table 3, we tested Med-VRAgent's reasoning strategy, hypothesizing that improved visual gudiance and feedback and higher-quality auxiliary information enhance performance. On the DeepSeek-VL-7B and MiniCPM-V2 models, Med-VRAgent outperformed others, achieving top scores in BVR (51.82 and 47.32) and average (46.74 and 43.41). Compared to Zero-shot, CoT, and ToT, it excelled in abnormality recognition, visual reasoning, and relational understanding, confirming the effectiveness

Methods	BLEU	ROUGE-L	METEOR
LLaVA-Med v1.5	9.64	12.26	8.21
Greedy	11.47	15.38	12.69
Beam Search	12.10	16.21	13.17
DoLa	11.79	15.82	12.72
OPERA	10.66	14.70	12.01
VCD	10.42	14.14	11.59
MedDr	12.37	16.45	13.50
FactMM-RAG	14.70	18.05	15.92
RULE	27.53	23.16	27.99
MMed-RAG	31.38	25.59	32.43
Med-VRAgent	**33.45**	**26.81**	**33.12**

Table 4: Comparison of Med-VRAgent with RAG methods, including FactMM-RAG, MMed-RAG etc, on the IU-Xray (BLEU, ROUGE-L, METEOR) dataset, based on **LLaVA-Med v1.5** model. The best score for each metric is highlighted in bold.

of the Med-VRAgent in complex medical VQA.

Performance Comparison of Decoding-based and RAG-based Methods. As shown in Table 4, on the IU-Xray dataset, LLaVA-Med v1.5 performed poorly (BLEU=9.64), with modest improvements from Greedy and Beam Search (BLEU=12.10). MMed-RAG showed significant improvement (BLEU=31.38), while Med-VRAgent achieved the best results (BLEU=33.45, ROUGE-L=26.81, METEOR=33.12), demonstrating that Med-VRAgent enhances medical report generation quality.

5 Discussion

This section presents three experiments examining Med-VRAgent's performance in medical visual reasoning tasks. The first investigates the importance of each component. The second explores the impact of MCTS width and depth on model accuracy. The third experiment evaluates the adaptive retrieval strategy (ARS) in the Reflection component compared with the traditional fixed Top-K method.

5.1 Analysis of Med-VRAgent's Components

We conduct an ablation study on **Med-VRAgent** to assess the contribution of its key components to medical visual reasoning. As shown in Fig 3, removing any component leads to performance degradation, highlighting the critical role of each module in reasoning progression, relevance, coherence, and adaptability. The visual extraction component has the greatest impact. Specifically, omitting any module increases the error rate in LLMs, affecting reasoning quality.

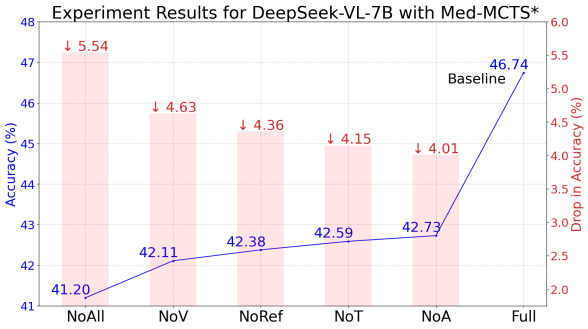

Figure 3: Ablation Experiment 1 Results (accuracy; %) for DeepSeek-VL-7B with Med-VRAgent on dataset GMAI-MMBench. Noall means removing all components, NoV means removing visual extraction, NOA means removing Assessor, and NoT means removing Teacher.

Experiment	Filter	Rerank	BLEU ↑	ROUGE-L ↑	METEOR ↑
Fixed Top-K	✗	✗	3.66	13.10	8.94
Rerank Only	✗	✓	3.75	13.20	9.22
Dynamic Top-K	✓	✗	3.80	13.75	9.12
Adaptive Retrieval	✓	✓	**3.90**	**14.10**	**9.58**

Table 5: Ablation study on the MIMIC-CXR dataset using the LLaVA-Med v1.5 model. Each retrieval strategy varies in its use of Filter and Rerank.

5.2 Width and Depth Optimization

We studied the impact of different MCTS fixed widths and depths on performance. The results are shown Fig 5. By adjusting the fixed width and depth in the search strategy, we found that accuracy could be improved. Search benefits decrease as width and depth rise, likely due to VLM's limited processing capacity. The best fixed combination (width 2, depth 3) achieved the highest accuracy of 46.68%. The adaptive strategy (width 1.74, depth 2.23) achieved an even higher accuracy of 46.74%. This result demonstrates that our adaptive strategy can maintain a balance between exploration and exploitation.

5.3 Evaluation of Adaptive Retrieval Strategy

The Table 5 presents an ablation study on the MIMIC-CXR dataset using the LLaVA-Med v1.5 model, evaluating different retrieval strategies. The experiments compare the impact of enabling filtering and Rerank mechanisms on the quality of generated outputs. The results indicate that using either Filter or Rerank alone leads to modest performance improvements. For instance, compared to the Fixed Top-K baseline, the Rerank Only strategy shows slight gains across all metrics . The best performance is Adaptive Retrieval, which combines both Filter and Rerank. It obtains the highest scores across all metrics.

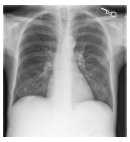	**Zero-shot:** ... clear lung field, ... The lungs appear to have normal bronchovascular markings, ... There is **no visible pleural effusion or pneumothorax, and the diaphragm appears intact**.... The lung parenchyma is free of significant pathological changes. The costophrenic angles are well-defined, and no blunting or fluid collection is visible. The chest X-ray appears normal .... **No signs of pneumonia, lung masses, or cardiovascular abnormalities are present.**	**GroundTruth:** ... The left lung is relatively well aerated and clear. The right hemithorax is markedly opacified with volume loss, circumferential pleural thickening and **pleural fluid** with near complete opacification of the right lung with right basal pleural catheter noted. ... Cardiac contours are somewhat obscured but unremarkable. ...Bibasilar opacities, larger on the left side, **could be due to atelectasis** but superimposed infection cannot be excluded. If any, there is a small right pleural effusion. There is elevation of the right hemidiaphragm. There is mild vascular congestion.	**Med-MAgent:** Increased density is observed in the left lower lung field, ... The grayscale value in this area is higher than the contralateral side, with reduced translucency, ... Blurred architecture of the left lower lobe.... ... **atelectasis ... should be considered**; ... Blurring of the right heart border may indicate involvement of the right middle or lower lobe, such as exudation or **pleural effusion**. Increased markings and reduced translucency ... possible inflammatory changes. ... No significant cardiomegaly or mediastinal shift is noted.

Figure 4: Med-VRAgent Medical Report Generation Case Study

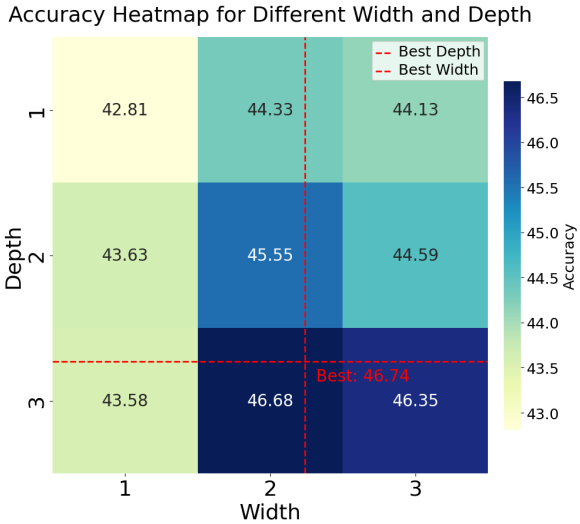

Figure 5: Ablation Experiment 2 Results (accuracy; %) for DeepSeek-VL-7B with Med-VRAgent on dataset GMAI-MMBench. Best is an adaptive exploration strategy, the average width and depth are 1.74 and 2.23 (red line), and other combinations are fixed width and depth.

5.4 Performance and Efficiency Analysis

In this experiment, we compared the performance of four methods (CoT, ToT, Med-VRAgent (Fix), Med-VRAgent (Ours)) on the GMAI-MMBench dataset. Fix is a fixed width of 2 and depth of 3. The results show that Med-VRAgent (Ours) performs best in terms of accuracy, reaching 46.74%. In addition, Med-VRAgent (Ours) has an advantage over Med-VRAgent (Fix) in inference time, which is 36.7 seconds, significantly lower than the fixed strategy of 45.7 seconds. Although the ToT method is slightly higher than CoT in accuracy (42.08% vs. 41.52%), its inference time is longer, reaching 31.3 seconds. The Cot method is the most efficient in inference time, only 18.3 seconds, but its accuracy is lower. Overall, Med-VRAgent (Ours) has achieved a good balance between accuracy and inference time, showing its comprehensive advantages over fixed strategies and other methods. This shows that adaptive strategies can optimize inference time while improving accuracy, and have better application potential.

Method	Accuracy (%)	Inference Time (s)
CoT	41.52	**18.3**
ToT	42.08	31.3
Med-VRAgent (Fix)	46.68	45.7
Med-VRAgent (Ours)	**46.74**	36.7

Table 6: DeepSeek-VL-7B compares the inference accuracy and average time of CoT, ToT and Med-VRAgent (fixed and adaptive policies) on the GMAI-MMBench dataset.

5.5 Case Study

As shown in the Fig 4, the case comes from the Deepseek-VL and the MIMI-CXR dataset. Med-VRAgent outperforms the Zero-shot in generating clinically accurate and factually grounded chest X-ray reports. While the Zero-shot model incorrectly states clear lungs and no pleural abnormalities, Med-VRAgent correctly identifies increased density, reduced translucency, and possible pleural effusion in the left lung, closely matching the expert GroundTruth. It avoids major hallucinations and captures subtle findings like blurred architecture and right heart border changes, suggesting infection or inflammation. Med-VRAgent also includes diagnostic considerations such as atelectasis, reflecting expert-level reasoning.

6 Conclusion

This study introduces Med-VRAgent, a novel medical visual reasoning framework that enhances multimodal large models' performance in medical image understanding. It incorporates a teacher-student-evaluator mechanism, visual guidance and self-feedback paradigm, and a multi-step reasoning strategy based on MCTS. Med-VRAgent achieved top performance across several medical multimodal benchmark datasets, demonstrating proficiency in image-text alignment, spatial structure understanding, and lesion recognition. Future research will focus on improving search efficiency, using advanced multimodal models, and expanding deployment in real clinical settings.

Limitations

Although Med-VRAgent has achieved significant improvements in medical visual reasoning, it still has limitations. Despite optimization, tree search is still resource-intensive. Due to node expansion strategies and computational resource constraints, it may not be possible to fully search all possible reasoning paths. It may not be directly transferable to other domains and additional domain adaptation is required. Visual guidance may have limited effect in complex images or low-quality images. Inaccurate reasoning may still occur when faced with fine-grained errors or very complex cases. Performance and reliability in actual clinical settings have not been fully verified.

Ethical Considerations

Ethical considerations are central to our research. In this study, we ensure adherence to ethical principles by exclusively using publicly available datasets and employing models that are open-source or widely accepted within the research community. We emphasize transparency in all stages of our work and prioritize the responsible application of technology, particularly in the sensitive domain of medical reasoning, to ensure that our contributions promote fairness, reliability, and societal benefit.

References

Jean-Baptiste Alayrac, Jeff Donahue, and Pauline Luc et al. 2022. Flamingo: a visual language model for few-shot learning.

Emily Alsentzer, John R. Murphy, and Willie Boag et al. 2019. Publicly available clinical bert embeddings.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

Oleg Boiko, Mahesan Seneviratne, et al. 2023. Emergent medical reasoning in large language models. *arXiv preprint arXiv:2305.17100*.

Tom B. Brown, Benjamin Mann, and Nick Ryder et al. 2020. Language models are few-shot learners.

Jiawei Chen, Dingkang Yang, Tong Wu, et al. 2025. Detecting and evaluating medical hallucinations in large vision language models. In *Proceedings of ICLR 2025*. Med-HallMark Benchmark.

Liang Chen, Ming Xu, and Yang Zhao. 2024a. Mmed-rag: Multi-modal retrieval-augmented generation for clinical decision support. *arXiv preprint arXiv:2401.12345*.

Pengcheng Chen, Jin Ye, Guoan Wang, and et al. 2024b. Gmai-mmbench: A comprehensive multimodal evaluation benchmark towards general medical ai. *arXiv preprint arXiv:2408.03361*.

Tianshu Cheng, Zongwei Zhou, et al. 2023. Segment anything model for medical images? *arXiv preprint arXiv:2304.04155*.

Aakanksha Chowdhery, Sharan Narang, and Jacob Devlin et al. 2022. Palm: Scaling language modeling with pathways.

Yung-Sung Chuang, Yujia Xie, Hongyin Luo, Yoon Kim, James Glass, and Pengcheng He. 2024. Dola: Decoding by contrasting layers improves factuality in large language models.

Dina Demner-Fushman, Marc D Kohli, Martin B Rosenman, and et al. 2015. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310.

Stefan Denner, Markus Bujotzek, and Dimitrios Bounias et al. 2025. Visual prompt engineering for vision language models in radiology.

Yunfan Gao, Yun Xiong, Xinyu Gao, and et al. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, Saizhuo Wang, Kun Zhang, Yuanzhuo Wang, Wen Gao, Lionel Ni, and Jian Guo. 2025. A survey on llm-as-a-judge.

Iryna Hartsock and Ghulam Rasool. 2024. Vision-language models for medical report generation and visual question answering: a review. *Frontiers in Artificial Intelligence*, 7.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep residual learning for image recognition.

Sunan He, Yuxiang Nie, Zhixuan Chen, Zhiyuan Cai, Hongmei Wang, Shu Yang, and Hao Chen. 2024. Meddr: Diagnosis-guided bootstrapping for large-scale medical vision-language learning. *arXiv e-prints*, pages arXiv–2404.

Lei Huang, Weijiang Yu, Weitao Ma, et al. 2023. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv preprint arXiv:2311.05232*.

Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, and Nenghai Yu. 2024. Opera: Alleviating hallucination in multi-modal large language models via over-trust penalty and retrospection-allocation.

Qidong Huang, Xiaoyi Dong, Pan Zhang, et al. 2025. Hallucination of multimodal large language models: A survey. *arXiv preprint arXiv:2404.18930*.

Zhecheng Ji, Nayeon Lee, Rita Frieske, and et al. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys (CSUR)*, 55(12):1–38.

Qiao Jin, Fangyuan Chen, Yiliang Zhou, et al. 2024. Hidden flaws behind expert-level accuracy of multimodal gpt-4 vision in medicine. *npj Digital Medicine*, 7:190.

Alistair EW Johnson, Tom J Pollard, Seth J Berkowitz, and et al. 2019. Mimic-cxr: A large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2017. Billion-scale similarity search with gpus. In *IEEE Transactions on Big Data*.

Yubin Kim, Hyewon Jeong, Shan Chen, et al. 2025. Medical hallucination in foundation models and their impact on healthcare. *medRxiv preprint*.

Joyce Lau, Swarnadeep Gayen, Dung Nguyen, and et al. 2020. A dataset of clinically generated visual questions and answers about radiology images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 781–791. Springer.

Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, and Lidong Bing. 2023. Mitigating object hallucinations in large vision-language models through visual contrastive decoding.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, and et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *arXiv preprint arXiv:2005.11401*.

Chunyuan Li, Cliff Wong, and Sheng Zhang et al. 2023a. Llava-med: Training a large language-and-vision assistant for biomedicine in one day.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023b. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models.

Yingshu Li, Yunyi Liu, Zhanyu Wang, et al. 2023c. A systematic evaluation of gpt-4v's multimodal capability for medical image analysis. *arXiv preprint arXiv:2310.20381*.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Jie Liu, Xin Huang, and Lin Zhao. 2023a. Rule: Retrieval-based unified learning for explainable medical vqa. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. 2024. Grounding dino: Marrying dino with grounded pre-training for open-set object detection.

Zhen Liu, Yichong Zhang, Denny Zhou, and et al. 2023b. Critique, suggest, and revise: An interactive framework for improving multistep reasoning. *arXiv preprint arXiv:2305.04388*.

Haoyu Lu, Wen Liu, and Bo Zhang et al. 2024. Deepseek-vl: Towards real-world vision-language understanding.

Aman Madaan, Niket Tandon, Prakhar Gupta, and et al. 2023. Self-refine: Iterative refinement with self-feedback. *arXiv preprint arXiv:2303.17651*.

Thomas Moell, Lin Zhang, Rui Chen, and Wei Xu. 2025. Medical reasoning in llms: An in-depth analysis of deepseek r1 on clinical tasks. *arXiv preprint arXiv:2404.09822*.

Ankit Pal, Logesh Kumar Umapathi, and Malaikannan Sankarasubbu. 2023. Med-halt: Medical domain hallucination test for large language models. In *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, pages 314–334.

Kishore Papineni, Salim Roukos, Todd Ward, and et al. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Alec Radford, Jong Wook Kim, and Chris Hallacy et al. 2021. Learning transferable visual models from natural language supervision.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks.

Daniel Rose, Vaishnavi Himakunthala, and Andy Ouyang et al. 2024. Visual chain of thought: Bridging logical gaps with multimodal infillings.

Hao Shao, Shengju Qian, Han Xiao, Guanglu Song, Zhuofan Zong, Letian Wang, Yu Liu, and Hongsheng Li. 2024. Visual cot: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning.

Tianxing Tang, Yifan Wang, Bolin Wang, and et al. 2023. Medagents: Large language models as medical collaborative agents for zero-shot medical reasoning. *arXiv preprint arXiv:2311.16592*.

Ryutaro Tanno, T. Barrett, David G. Andrew Sellergren, et al. 2025. Collaboration between clinicians and vision–language models in radiology report generation. *Nature Medicine*, 31:599–608.

Hugo Touvron, Louis Martin, and Kevin Stone et al. 2023. Llama 2: Open foundation and fine-tuned chat models.

Zhuoyuan Wang, Hao Zhang, et al. 2023. Hierarchical visual prompt tuning for medical image classification. *Medical Image Analysis*, 87:102845.

Jason Wei, Xuezhi Wang, and Dale Schuurmans et al. 2023. Chain-of-thought prompting elicits reasoning in large language models.

Jason Wei, Xuezhi Wang, Dale Schuurmans, and et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*.

Liyuan Wu, Xiaoyu Zhang, Yichong Huang, and et al. 2025. Medreason: Structured reasoning paths for faithful medical problem solving. *arXiv preprint arXiv:2402.05727*.

Peng Xia, Kangyu Zhu, and Haoran Li et al. 2025. Mmed-rag: Versatile multimodal rag system for medical vision language models.

Yuxi Xie, Kenji Kawaguchi, Yiran Zhao, Xu Zhao, Min-Yen Kan, Junxian He, and Qizhe Xie. 2023. Self-evaluation guided beam search for reasoning.

Yuta Yamamoto, Tyna Eloundou, et al. 2023. Can large language models reason about medical questions? *arXiv preprint arXiv:2305.17594*.

Shinn Yao, Jeffrey Zhao, and Dian Yu et al. 2022. React: Synergizing reasoning and acting in language models.

Shunyu Yao, Dian Yu, and Jeffrey Zhao et al. 2023. Tree of thoughts: Deliberate problem solving with large language models.

Yuan Yao, Tianyu Yu, and Ao Zhang et al. 2024. Minicpm-v: A gpt-4v level mllm on your phone.

Tong Yu, Yukun Huang, et al. 2023. A comprehensive benchmark of vision-language models for medical imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21127–21137.

Wenhao Yu, Ji He, Qingyang Yin, and et al. 2025. Finemedlm-o1: A finetuned medical llm for long-form clinical reasoning and test-time adaptation. *arXiv preprint arXiv:2403.04042*.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Xian Li, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. 2025. Self-rewarding language models.

Hao Zhang, Jun Ma, Hongyu Wang, et al. 2023a. Large multimodal models can be easily distracted by spurious correlations in medical imaging. *arXiv preprint arXiv:2307.16311*.

Wei Zhang, Rui Wang, and Yuxuan Li. 2023b. Factmm-rag: Factual multi-modal retrieval-augmented generation for medical vqa. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Wenhui Zhang, Jianlong Fu, and Yijuan Lu et al. 2023c. Reinforced ranker-reader for open-domain question answering.

Denny Zhou, Nathanael Schärli, Le Hou, and et al. 2022. Least-to-most prompting enables complex reasoning in large language models. *arXiv preprint arXiv:2205.10625*.

Deyao Zhu, Jun Chen, and Xiaoqian Shen et al. 2023. Minigpt-4: Enhancing vision-language understanding with advanced large language models.

Kangyu Zhu, Ziyuan Qin, Huahui Yi, Zekun Jiang, Qicheng Lao, Shaoting Zhang, and Kang Li. 2025. Guiding medical vision-language models with explicit visual prompts: Framework design and comprehensive exploration of prompt variations.

Kangyu Zhu, Peng Xia, and Yun Li et al. 2024a. Mmedpo: Aligning medical vision-language models with clinical-aware multimodal preference optimization.

Zhiying Zhu, Yiming Yang, and Zhiqing Sun. 2024b. Halueval-wild: Evaluating hallucinations of language models in the wild. *arXiv preprint arXiv:2403.04307*.

A Prompt

Prompt

Role Setting: You are a medical expert providing guidance on medical image analysis to help students improve their understanding.
Task Description: Focus on the red-boxed area in the image, using previous guidance and student feedback to offer optimized suggestions for enhancing their analysis skills.
Guidance Content:
Analyze Key Area:
Identify the red-boxed region for closer analysis.
Observe structural features, shape changes, color contrasts, and any abnormalities.
Reference Feedback and Suggestions:
Evaluate the student's previous analysis.
Point out missed details or inadequate analysis, and offer visual techniques.
Optimize Analysis Directions:
Guide the student based on the image type (e.g., CT, X-ray, ultrasound).
Suggest perspectives like cross-sections or tissue density changes.
Important Notes:
Your goal is to help students master image analysis, not to do it for them.
Focus on a logical, systematic approach for comprehensive image interpretation.
Previous Guidance: </Guidance>
Student's Answer: </Answer>
Feedback Information: </Feedback >
Use this format for guidance: </Guidance> Guidance here </Guidance>

Figure 6: Evaluation prompt for medical image analysis results

Prompt

You are a medical expert. Please review the image and visual analysis guidance and rate the student-generated answers using the additional 5-point rating system described below. The rating will be cumulative based on the following criteria:
5-point rating scale:
1. Relevant information: If the medical vision answer provides some information that is relevant to the user's query, even if the information is incomplete or contains some incompletely relevant content, 1 point can be awarded. 2. Partially solve the problem: If the answer solves most of the user's question, but does not fully answer the user's question or does not directly answer the core query, 2 points can be awarded. 3. Essential elements: If the answer answers the basic elements of the user's question from a medical vision perspective, although it may lack detail or completeness in some aspects, but is still helpful to the user, 3 points can be awarded. 4. Direct and comprehensive solution to the problem: If the answer directly and comprehensively solves the user's question, although there may be some room for improvement in clarity, conciseness or visual focus, 4 points can be awarded. 5. Tailored, professional and profound: If the answer is tailored to the user's question, provides an in-depth and professional answer through medical vision, avoids irrelevant information, and produces high-quality, engaging and insightful content, 5 points should be awarded.
Information: <guidance> Teacher's guidance </guidance> <answer> Student's answer </answer>
Evaluation steps:
Total rating: Please briefly explain your rating in 100 words or less.
Suggestions for teachers: Provide suggestions for teachers to build better guidance in 100 words or less.
Revision suggestions for students: Provide revision suggestions for students in 100 words or less.
Rating conclusion:
<score>Integer score</score>
<feedback1>Feedback to teachers</feedback1>
<feedback2>Revision suggestions for students</feedback2>

Figure 7: Evaluation prompt for medical image analysis results